

Iria da Cunha Fanego

POR JOSÉ MANUEL POSADA DE LA CONCHA

A escasos 25 km al sur de Santiago de Compostela, en Galicia, cerca de la frontera con Portugal, se encuentra el pueblo de *Iria Flavia*, antigua ciudad galaica, que según la tradición tomó el nombre de una mujer que predicó al lado del Apóstol Santiago. En esta parte de España se habla una de las cuatro lenguas romances de la península ibérica (las otras son el catalán, el español y el portugués). Aunque Galicia es una provincia de España, su idioma es más cercano al portugués que a la lengua de Cervantes.

De los muchos apellidos castellanizados cuyo origen es la lengua lusa, hay algunos verdaderamente memorables, como Borges; otros muy extendidos como Yáñez, Frago o Figueroa. En particular, el apellido da Cunha, se asentó en Galicia antes de castellanizarse como Acuña (más diseminado en México que en España) y su **nh** se pronuncia igual que la muy española **ñ**. *Iria da Cunha* es el nombre de una investigadora gallega a quien sus padres no llamaron así por el personaje que dicen acompañaba al apóstol, sino por lo bello del pueblo cercano a Portugal, en la orilla de un camino muy transitado. Con su segundo apellido, Fanego —que no más de 500 personas comparten en España, la mayoría en La Coruña—, resulta que Iria da Cunha Fanego es tan gallega como el *pulpo á feira*¹.

De esto charla con nosotros la doctora en la antesala de su entrevista. Sus conocimientos no están generados sólo por la curiosidad de conocer sobre el origen de sí misma, sino porque es filóloga hispánica de carrera pero, como es de esperarse, de cultura gallega desde la cuna.

Ya entrados en el tema sobre sus proyectos, comenta:

Tuve la oportunidad de hacer una estancia en la Universidad de São Paulo en Brasil, ya que allí realizan investigaciones en ingeniería lingüística en líneas muy similares a las que yo desarrollo. Me interesó estar con ellos, entre otras cosas, por la cercanía lingüística del español con el portugués que, por mi origen gallego, puedo entender sin mayor problema.



La Dra Iria “sosteniendo” un gran corpus lingüístico: la Biblioteca Central

Me di cuenta de que muchos desarrollos de los brasileños pueden ser adaptados con relativa facilidad al español. En particular, el análisis discursivo automático, que tiene que ver directamente con la traducción automática de un texto de un idioma a otro, por medio de software. Actualmente, la mayoría de los sistemas de traducción automática emplean estrategias estadísticas, que detectan cuál es la traducción más probable de cada oración dado un corpus de grandes dimensiones en varios idiomas. Estos sistemas funcionan, pero las traducciones que ofrecen todavía distan mucho de una traducción humana. Por eso deben ser optimizados con métodos lingüísticos. Si, por ejemplo, realizamos un análisis discursivo del texto que queremos traducir para obtener un árbol argumentativo que exprese sus ideas y saber cómo están relacionadas unas con otras (ver ejemplo 1, página 13), ese análisis ayuda muchísimo; Esta es la esencia lingüística de mi trabajo. Entre las frases que componen un texto, vamos a encontrar relaciones de causa, de resultado, de condición, de lista, de contraste, etc. Si en español tenemos una herramienta de análisis discursivo que nos proporcione su estructura y ésa es, justamente, la estructura que le damos en el idioma al que traducimos, el resultado será más cercano al texto producido por el autor en el idioma original (ver ejemplo 2.1, página 13). Las herramientas de análisis discursivo ya existen para el inglés, el japonés y el portugués de Brasil. Nosotros vamos por buen camino para el español. En el futuro prevemos adaptarlo para el francés, que es otra lengua romance, y un grupo de investigación del País Vasco tiene en mente adaptarlo para el euskera, que no es romance pero es otra de las lenguas oficiales de España.

Después de estudiar filología hispánica en Santiago de Compostela, la doctora da Cunha realizó una maestría y un

doctorado en lingüística aplicada en la Universitat Pompeu Fabra de Barcelona, en lengua catalana. Posteriormente, durante una estancia de investigación en el Laboratorio Informático de Avignon (Francia) conoció al doctor Gerardo Sierra y de ahí comenzó una estrecha relación académica y laboral, que se tradujo en esta estancia posdoctoral de un año, que acaba de comenzar, en el Grupo de Ingeniería Lingüística (GIL) del Instituto de Ingeniería.

Otro uso del análisis discursivo automático para el español –continúa Iria da Cunha– es didáctico, como herramienta para profesores de asignatura de letras, particularmente en análisis del discurso, para explicar en qué consiste éste con ayuda de árboles discursivos. Justamente yo lo emplearé durante este año al impartir tres materias en la UNAM: Análisis del Discurso II (en la Licenciatura en Lengua y Literaturas Hispánicas), Minería de Textos (en la Licenciatura en Ingeniería en Computación) y Tratamiento de Lenguaje Natural (en el Posgrado de Ciencia e Ingeniería en Computación).

*Otro uso, igual de práctico aunque represente un reto considerablemente mayor, es la obtención de resúmenes automáticos de textos. Esencialmente lo que hacemos para obtener un resumen automático es identificar los elementos más importantes del texto y separarlos de los menos relevantes. Hoy en día, los resúmenes son importantísimos, pues en la sociedad de la información hay una explosión de documentos que prácticamente no podemos atender como se debe al realizar una investigación sobre un tema dado. No es mentira si decimos que para un término, por ejemplo **gen**, encontramos varios cientos de millones de artículos en un buscador como Google. Tener un resumen de cada artículo, sin que el propio texto lo proporcione, sino a través de programas como los que trabajamos en ingeniería lingüística, es una ayuda fundamental para aceptar o desechar información. Los resúmenes automáticos ahorran mucho tiempo y esfuerzo. Así, si obtenemos automáticamente un árbol discursivo de un texto, podemos seleccionar las frases que tienen más relevancia en dicho árbol. Por ejemplo, si extraemos el elemento principal del árbol del ejemplo 2.2 (la frase 4), tendremos un buen resumen del texto, ya que precisamente está expresando el objetivo del estudio que expone.*

Haciendo una pausa en la plática sobre sus proyectos, nos expone las razones por las que se encuentra en la UNAM.

Esta universidad tiene bastante prestigio en Europa, particularmente en España, pues, entre muchas otras cosas, el año pasado le otorgaron el Premio Príncipe de Asturias y todo mundo se enteró. Por esta razón, cuando salieron las becas posdoctorales de la UNAM, no lo pensé dos veces y me presenté como candidata. También estoy aquí por otra razón fundamental: en el futuro me gustaría ser docente e investigadora en alguna universidad española y uno de los requisitos para ello es haber realizado estancias en grupos de investigación reconocidos en todo el mundo. Por eso me encuentro ahora en el grupo del doctor Sierra, un grupo con muchos proyectos interesantes y publicaciones muy reconocidas.

Retomando sus proyectos de investigación, apunta:

Otra aplicación del análisis discursivo en la que trabajamos, un poco más compleja, es la evaluación automática de resúmenes, que es un reto bastante interesante. ¿Cuál es el mejor resumen entre 20, sobre un mismo texto, por ejemplo?

En todo el mundo hay concursos de robots entre universidades o empresas, y también de software para jugar ajedrez o cosas similares. Pero pocos saben que existen concursos que premian al grupo de ingeniería lingüística que obtenga el mejor resumen de un texto dado. Aunque ahora hay que verlo desde los ojos del juez: se obtienen los resúmenes automáticamente, pero, ¿cómo los evaluamos para saber cuál es el mejor? Hace unos 15 años, la evaluación se hacía de forma manual, los especialistas leían todos los resúmenes y dictaminaban. Esto siempre implicó mucha inversión en personas y en tiempo. Luego se pensó en hacer sistemas semiautomáticos. Se pedía a gente que redactase resúmenes sobre un texto y se comparaban ambos resúmenes, el de las personas y el de las máquinas. Fue un paso adelante, pero también requería fuerte inversión. Hace un año o dos se abrió una nueva línea de investigación en EUA, que es la evaluación totalmente automática de los resúmenes, donde se aplican técnicas estadísticas; es donde ahora nos encontramos nosotros también y ya estamos obteniendo resultados prometedores.

Todas las mañanas, al salir de mi casa en Copilco el Alto, paso por una bolsa de fruta con limón y picante –me encanta–. He oído de “la venganza de Moctezuma”, pero por suerte a mí no me ha pasado, todavía-. La doctora comenta lo anterior mientras degusta unas enchiladas de mole negro oaxaqueño en el restaurante Azul y Oro y se ufana de haber comido unos riquísimos chapulines rojos el fin de semana pasado. –Poco a poco iré conociendo el país. Es la primera vez que estoy aquí, pero será por un año entero, así que no hay prisa por probar los gusanos de maguey con un tequila o unos chiles en nogada. Poco a poco–.

Para finalizar la charla, le hacemos la pregunta expresa: ¿Para el desarrollo óptimo de sus proyectos de investigación es necesario dominar varios idiomas?

Para realizar sistemas de ingeniería lingüística existen dos tipos de estrategias principales: la estadística y la lingüística. Si se sigue la primera estrategia se toman los textos como conjuntos de palabras y se utilizan como matrices de datos sobre los que se aplican técnicas puramente estadísticas. Estas técnicas suelen ser independientes del idioma. Si se sigue la segunda estrategia se tratan los textos desde un punto de vista lingüístico, y se analizan, por ejemplo, sus unidades léxicas, su estructura sintáctica, su estructura discursiva, la semántica, etc. Todos estos aspectos sí dependen, por lo general, de la lengua con la que se esté trabajando (las palabras de cada lengua son diferentes y las estructuras sintácticas del inglés y el español, por ejemplo, también lo son). Es aquí donde saber idiomas entra en juego. Yo estoy desarrollando un sistema, un analizador discursivo automático para el español, pero estoy adaptando este sistema de uno ya existente para el portugués, y evidentemente ha sido de gran

ayuda conocer esta lengua. Tenemos previsto en el futuro adaptar el sistema al francés, y no lo podría hacer si no supiese francés. La adaptación al chino no la veo factible porque desconozco esta lengua; necesitaría contar con colaboradores para llevarla a cabo. En conclusión: saber lenguas es necesario para realizar la adaptación de los sistemas a otras lenguas y además le sirve al investigador para ser autónomo a la hora de hacerlo.

Con esto nos despedimos de ella, mientras caminamos hacia la Biblioteca Central para tomarle una fotografía. 📷

Para más información sobre los proyectos o publicaciones de Iria da Cunha pueden acudir a su página web personal: <http://www.upf.edu/pdi/iula/iria.dacunha/>

¹ Plato típico gallego.

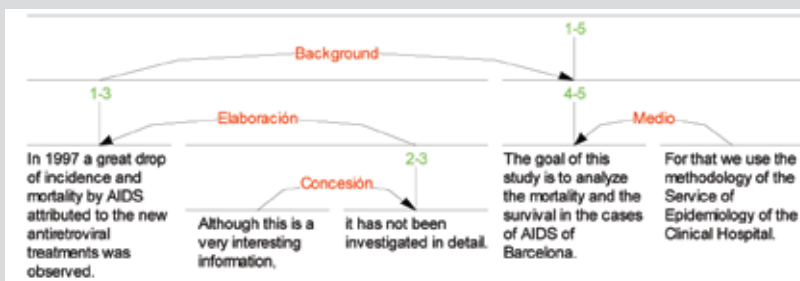
1) ANÁLISIS DISCURSIVO DE UN FRAGMENTO DE UN ARTÍCULO MÉDICO

“En 1997 se observó un gran descenso de la incidencia y la mortalidad por sida atribuida a los nuevos tratamientos antirretrovirales. Aunque este es un dato de gran interés, no se ha investigado con detalle. El objetivo de este estudio es analizar la mortalidad y la supervivencia en los casos de sida de Barcelona. Para ello empleamos la metodología del Servicio de Epidemiología del Hospital Clínico.”



2) EJEMPLOS DE APLICACIONES DEL ANÁLISIS DISCURSIVO

2.1. TRADUCCIÓN AUTOMÁTICA. Una buena traducción español-inglés de este fragmento debería incluir la misma estructura discursiva en las dos lenguas:



2.2. RESUMEN AUTOMÁTICO. Si seleccionamos los elementos más importantes (“núcleos”) y eliminamos los menos importantes (“satélites”), obtendríamos el siguiente resumen:

